

- Signal peptide exchange for higher expression of *Bjerkandera adusta* Loosenin 1 in *P. pastoris*.

Linus Brinkenstråhle



Degree project • 15 credits

Biology with specialisation in Biotechnology

Molecular Sciences, 2018:13

Uppsala, 2018

–Signal peptide exchange for higher expression of
Bjerkandera adusta Loosenin 1 in *P. pastoris*.

Linus Brinkensträhle

Supervisor: Jerry Ståhlberg, Swedish University of Agricultural Sciences,
Department of Molecular Sciences
Assistant supervisor: Topi Haataja, Swedish University of Agricultural Sciences,
Department of Molecular Sciences
Examiner: Mats Sandgren, Swedish University of Agricultural Sciences,
Department of Molecular Sciences

Credits: 15 credits
Level: First cycle, G2E
Course title: Independent project in Biology
Course code: EX0689
Programme/education: Biology with specialisation in Biotechnology

Place of publication: Uppsala
Year of publication: 2018
Cover picture: Linus Brinkensträhle
Title of series: Molecular Sciences
Part number: 2018:13
Online publication: <https://stud.epsilon.slu.se>

Keywords: Loosenin, cellulose, signal peptide, expansin, *Pichia pastoris*,
recombinant expression

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences (NJ)
Department of Molecular Sciences

Abstract

The protein Loosenin 1 from the wood degrading fungus *Bjerkandera adusta* (BaLOOS1) has been shown to loosen up recalcitrant biomass and enhance enzymatic saccharification. BaLOOS1 is the only loosenin protein that has been biochemically characterised so far. An earlier attempt to express BaLOOS1 in *Pichia pastoris*, with its native signal peptide (SP), indicated low expression level (T. Haataja and J. Ståhlberg, unpublished).

The purpose of this study was to exchange the native SP of BaLOOS1 in an attempt to obtain higher expression levels. A pGAP *P. pastoris* expression plasmid was used that contained the gene for BaLOOS1 with an appended C-terminal His tag, and the native SP was exchanged by Gibson assembly followed by transformation into *P. pastoris* X33-strain. Four SPs were tested. Two SPs were from loosenin homologs in *Trichoderma reesei* and *P. pastoris*, respectively, from a previous study. The two other SPs were the α -factor from *Saccharomyces cerevisiae* and the Epx1-SP from the most abundant secretory protein Epx1 in *P. pastoris*. All the SP constructs were successfully inserted. The *T. reesei* SP gave the strongest band of the expected molecular weight on SDS-PAGE gel of the tested SPs indicating a higher expression level than the others. Specific binding to an IMAC column strongly suggests that the expressed protein is the His-tagged BaLOOS1. However, definite confirmation of the identity remains to be done, e.g. by peptide mapping.

Keywords: Loosenin, cellulose, signal peptide, expansin, *Pichia pastoris*, recombinant expression

Table of contents

Abbreviations	5
1 Introduction	6
2 Materials and methods	9
2.1 Design of gene constructs for expression	9
2.2 Gene assembly, DNA preparation and analysis	9
2.3 Transformation of <i>E. coli</i> and plasmid preparation	10
2.4 Transformation of <i>P. pastoris</i> and colony PCR analysis	10
2.5 Protein expression	11
2.6 Estimation of protein expression levels	11
3 Results	13
3.1 Design of gene constructs for expression	13
3.2 Gene assembly, DNA preparation and analysis	15
3.3 Transformation of <i>E. coli</i> and plasmid preparation	16
3.4 Transformation of <i>P. pastoris</i> and colony PCR analysis	17
3.5 Protein expression	18
4 Discussion	21
References	23
Acknowledgements	25

Abbreviations

ADC, aspartate- α -decarboxylase
BaLOOS1, *Bjerkandera adusta* loosenin
cPCR, colony PCR
DPBB, double psi beta barrel
EGV, endoglucanase V
ER, endoplasmic reticulum
GH45, glycoside hydrolase of family 45
Ig, immunoglobulin
KpLOOL, *P. pastoris* loosenin-like protein
LiOAc, lithium acetate
MltA, murein lytic transglycolase A
PDB, protein data bank
SP, signal peptide
SRP, signal recognition particle
TrCP, *Trichoderma reesei* cerato-platanin
TrLOOL, *Trichoderma reesei* loosenin-like protein

1 Introduction

The most abundant carbon source in the world is in the form of different plant cell wall materials. Which have a function to resist and protect the cell from the harsh surroundings e.g. cellulose and lignin. This makes the structure of these materials sturdy and resistant towards most harsh conditions the plant gets exposed to. Thus, degradation for industrial use of these materials is usually highly energy demanding and requires several different treatments (Dal Piccoli, Aver et al. 2018, Martinez-Gutierrez 2018). This leads to expensive treatments of chemical, enzymatic and mechanical nature to separate the polymers so that the simpler carbon structure can be used for e.g. biofuel (Galicia-Medina, Barrios-Estrada et al. 2018). Therefore, effectiveness of plant material degradation is always sought after and needed.

Scientists have therefore made an effort to make use of the multitude of organisms that degrade plant materials. Nature's own plant degraders, the fungal community, play an important role in the global carbon recycling with many different species adapted to parasitic as well as saprotrophic life styles. Many fungi have the ability to degrade the sturdy materials to obtain the otherwise unavailable sugars of the plants. For this purpose, filamentous fungi are the most suited with white-rot fungi of the phylum Basidiomycota being an especially efficient group (Wang, Vazquez-Duhalt et al. 2003, Suzuki, Vuong et al. 2014, Lopez, Peng et al. 2018, Loyd, Held et al. 2018). The main focus of this study is a protein from the white-rot fungus *Bjerkandera adusta*, which has shown great promise degrading lignin (Wang, Vazquez-Duhalt et al. 2003). The protein of interest is the *Bjerkandera adusta* Loosenin (BaLOOS1), which has been reported to bind to both cellulose and chitin, disrupt cotton fibres and boost the enzymatic saccharification of recalcitrant cellulosic biomass (Quiroz-Castaneda, Martinez-Anaya et al. 2011).

Loosenins show homology with expansins and glycoside hydrolases of family 45 (GH45). The structure homology is found in the double psi beta barrel (DPBB)

domain which is a common structural motif shared with several other protein super families e.g. murein lytic transglycolases (MltA), aspartate- α -decarboxylase (ADC) and endoglucanase V (EGV) (Castillo, Mizuguchi et al. 1999, Payne, Knott et al. 2015). However, the BaLOOS1 is only 109 amino acids in length with only one domain, which means that it lacks the additional immunoglobulin (Ig) domain present in expansins. The Ig domain extends the presumed substrate-binding surface and is thus believed to help these proteins to bind with a larger area to cell wall components. Expansins are found in all plants and are known to be responsible for loosening up of the cell wall to allow cell expansion and growth, without seriously compromising its structural integrity (McQueenmason and Cosgrove 1994). As with BaLOOS1, expansins have not shown any hydrolytic activity and only seem to work in a disruptive way. However, it has been shown that BaLOOS1 can boost the effect of cellulase enzymes. This was done with *Agave tequilana* bagasse, which was first treated with BaLOOS1, and then by a cocktail of different cellulases and xylanases. The pre-treated bagasse showed a hydrolytic effect 7.5 times higher than the bagasse without pre-treatment. Although BaLOOS1 is the only loosenin studied so far, it has 200 BLAST hits with 40% sequence identity in fungi which indicates that loosenin is a ubiquitous protein in the plant degrading community of the fungal world. This also means that it is an important protein to study and make use of.

An unpublished pilot study performed by T. Haataja and J. Ståhlberg had problems expressing the BaLOOS1 in *P. pastoris*, whereas the other homologous proteins were successfully expressed (Figure 1). However, since BaLOOS1 is the only one of those proteins with a proven plant material disrupting ability so far, it is the primary choice for further catalytic and structural studies. Therefore, an attempt to increase the expression level of the BaLOOS1 in *P. pastoris* is made in this study. In recent years *P. pastoris* has been assigned other genus and species names and is currently also known as *Komagataella pastoris* or *K. phaffi*. For simplicity the name *P. pastoris* will most often be used, in this thesis.

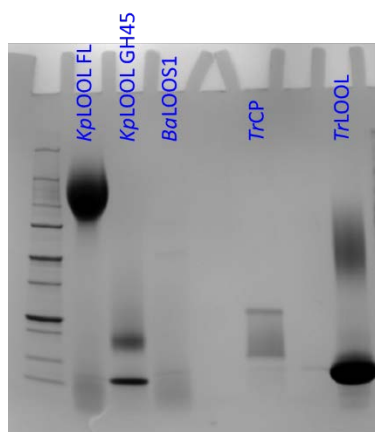


Figure 1: SDS-PAGE analysis of the expression of loosenin homologs in *P. pastoris* in the previous pilot study, with the *P. pastoris* loosenin-like protein in lane 2-3 with the full-length protein in lane 2 and only the GH45-domain in lane 3. In lane 4 is the *Bjerkandera adusta* loosenin. In lane 6 is the *Trichoderma reesei* cerato-platanin and in lane 8 is the *Trichoderma* loosenin-like protein.

In all secretory proteins, like the BaLOOS1, there is a short sequence in the beginning of the unmaturing protein. The sequence is the first being translated and folds outside the ribosome to attach to the signal recognition particle (SRP) recognition site which pauses the translation of the protein. The ribosome then travels through the cytoplasm and attaches to the endoplasmic reticulum (ER) translocon by binding to the SRP receptor. Once inside the ER lumen the SP gets cleaved off by a signal peptidase (SPase). This sequence is called a signal peptide and usually consists of 16-30 amino acids with a positively charged N-region. This is followed by a hydrophobic central h-region of about 5-15 amino acids and then a c-region where the cleavage site for the SPase occurs (Kapp, Schrempf et al. 2013). Since the signal peptide is a crucial part of the expression of secretory proteins the idea of this study was to exchange the native signal peptide of BaLOOS1, which would hopefully increase the expression of the BaLOOS1 enough for crystallization and catalytic studies.

2 Materials and methods

2.1 Design of gene constructs for expression

The web-based SignalP program (Nielsen 2017) was used to make a signal peptide prediction to visualise the cleavage site of the signal peptide and the N-terminus of BaLOOS1. The sequence without the signal peptide was then used to find homologous proteins using ncbi BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the Protein data bank database (PDB) (Berman, Westbrook et al. 2000). The PDB-files from the protein homologs were used together with a model of the BaLOOS1 created with the Phyre2 software (Kelley, Mezulis et al. 2015) to superimpose the models onto each other using the Pymol software (Schrodinger 2015) to study the structure of the N-terminus. To acquire new signal peptide constructs a signal peptide prediction was conducted. The prediction was performed using the sequences from the pilot study, kindly supplied by T. Haataja, and other signal peptides of interest. An alignment was made using the BaLOOS1 amino acid sequence and the sequences of interest in the program MEGA7 (Kumar, Stecher et al. 2016) with the use of the MUSCLE algorithm (Edgar 2004, Edgar 2004) and default settings. To insert the constructs into a plasmid the SnapGene software (from GSL Biotech; available at snapgene.com) was used, where the new signal peptides were incorporated into the BaLOOS1 protein inside a pMA_Hi6108_NoHis_Zeo plasmid. A Gibson assembly overlap and a Primer (Primer 1: 5'-ACTG-GATTGCAACACGG-3') was also constructed, see Figure 2.

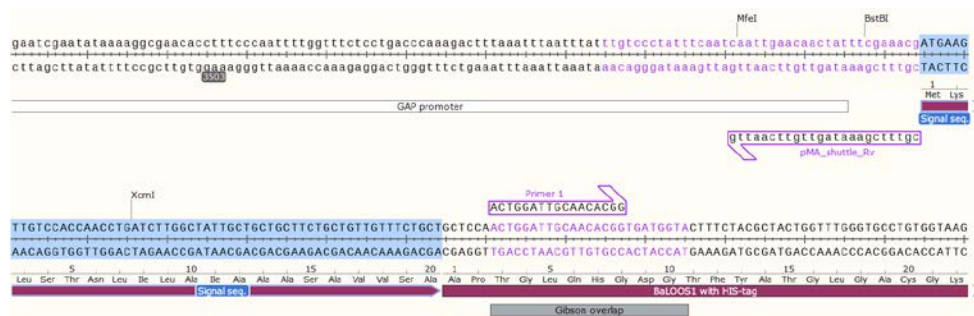


Figure 2: The picture shows a screenshot of the Snapgene software with the signal peptide from the *Epx1-SP* construct, marked in blue, together with the downstream Primer 1 and the Gibson overlap which are a part of the BaLOOS1 mature protein sequence.

2.2 Gene assembly, DNA preparation and analysis

All synthetic oligonucleotides used in this study are listed in Appendix 1. The signal peptide DNA constructs ordered from New England Biolabs arrived dry and nuclease free. Water was added to obtain a concentration of 0.05 pmol/μl. This was also done with Primer 1 which got diluted to a concentration of 10 μM. The purity of the constructs was analysed by a 1% agarose gel containing GelRed from Biotinum. *E. coli* Top10 cells containing the plasmid with the gene for BaLOOS1

with an appended C-terminal His₆ tag (AAASHHHHHH) were kindly provided by T. Haataja and were cultured in standard LB+amp media with 100 µg/ml of ampicillin at 37°C and 130 rpm for 16 h. To verify the amount of BaLOOS1 plasmid within the Top10 cells a plasmid preparation was done using Thermo Scientific GeneJET Plasmid Miniprep Kit according to supplier's instructions. A backbone PCR following New England Biolabs PCR using Q5® High-fidelity DNA polymerase 2X Master Mix with a forward primer (Primer 1: 5'- ACTG-GATTGCAACACGG-3') and a reverse primer (pMA_Shuttle_Rv: 5'-CGTTTCGAAATAGTTGTTCAATTG-3') with an extension time of 20 sec at 61°C was performed. It was purified by gel electrophoresis using a 1% agarose gel and extracted with the Thermo Scientific GeneJET Gel Extraction Kit. A Gibson assembly was conducted using New England Biolabs NEBuilder® HiFi DNA Assembly Master Mix/NEBuilder HiFi Assembly Cloning Kit according to the supplied instructions.

2.3 Transformation of *E. coli* and plasmid preparation

The transformation was performed with *E. coli* Top10 competent cells following the One Shot® TOP10 Competent Cells protocol from Invitrogen instead of the suggested strain in the Gibson assembly protocol from New England Biolabs. The cells were then grown on standard LB+amp plates, containing 200 µg/ml and 100 µg/ml ampicillin, at 37°C. To make sure that the plasmid had been taken up by the Top10 cells a colony PCR (cPCR) was conducted with 5 µl Green Taq 2x master mix, 0,5 µl Forward primer (Sequencing_Fw: 5'-CAGCCCAGGGATGGAAAAG-3'), 0,5 µl reverse primer (pGAP_LPMO_rev: 5'-GCAAATGG-CATTCTGACATCC-3'), 1 µl *E. coli* DNA template, 3 µl water. Miniprep plasmid preparations were performed to examine the DNA concentration, following the same protocol as aforementioned.

2.4 Transformation of *P. pastoris* and colony PCR analysis

A wild type culture of *Pichia pastoris* X33-strain was grown in 5 ml standard YPD-media at 30°C and 120 rpm until reaching an OD₆₀₀ of 2.0. They were then transferred to make a larger, 50 ml culture that was grown in the same conditions until reaching an OD₆₀₀ of 1.0 and then got diluted with more YPD to OD₆₀₀ of 0.8. The cultures were then transferred to 50ml falcon tubes and centrifuged at 500 xg for 5 min, discarding the supernatant afterwards. The plasmid linearization experiment used 4 µl Sfi, 4 µl 10x HD buffer, 8 µg DNA template and nuclease free water to get a final volume of 40 µl. The mixture was heated at 50°C for 3 h followed by a PCR clean-up using Thermo Scientific GeneJET PCR clean up kit, and the DNA was eluted with 20 µl of water. To make the cells competent, 9 ml BEDS and 1 ml 1 M DTT was added to the Falcon tubes and they were put on a shaker at 100 rpm and 30°C for 5 min. The tubes were then centrifuged for 5 min at 500 xg, the supernatant was discarded. The pellet was resuspended in 500 µl of BEDS and then put on ice for 2 min. For the transformation, electroporation was used with

the settings 1500 V, 200 μ s, 50 μ F and 9 ms. After the electroporation 1 ml YPD preheated to 30°C was added to the electroporation cuvette and incubated on a shaker for 1 h at 100 rpm and 30°C. The cells were then spread on YPD+zeo plates preheated to 30°C in two different ways with one undiluted and one diluted. The dilution used 950 μ l of 0.9 % NaCl solution and 50 μ l cell solution. The spread was done with 100 μ l of the diluted cells being spread to the plate while the undiluted was spread with 50 μ l cell solution. The plates from both spreading's were then incubated at 30°C for 2 days.

Three colonies from each construct was picked for cPCR and re-spreading. The cPCR was performed by suspending the colonies in a 100 μ l of 200 mM lithium acetate (LiOAc) and 1 % SDS solution. The solution was vortexed and incubated for 5 min in a 70°C heating block. To precipitate the DNA 300 μ l of 99 % ethanol (EtOH) was added and the solution was mixed by vortexing and then centrifugated at 15000 xg for 3 min to collect the DNA. The supernatant was discarded, and the pellet was washed with 500 μ l of 70% EtOH. To suspend the pellet 100 μ l of water was added and the samples were centrifugated for 1 min at 15000 xg to remove debris. The PCR solution was mixed with 5.5 μ l of Green Taq master mix, 0.5 μ l of forward primer (Sequencing_LPMO_Fw: 5' - CAGCCCAGGGATGGAAAAG-3'), 0.5 μ l of reverse primer (pGAP_LPMO: 5' - GCAAATGG-CATTCTGACATCC-3'), 2 μ l of DNA template and 2.5 μ l water. The PCR program was set as follows initial denaturation 95°C for 2 min, 35 cycles; 95°C for 45 secs, 54°C for 45 secs and 72°C for 2 min this was followed by a final extension at 72°C for 7 min and a hold at 4°C for infinity. The PCR products were then analysed by absorbance measurement with a Thermo Scientific Nanodrop and a gel electrophoresis with a 1% agarose gel containing GelRed.

2.5 Protein expression

For each of the five constructs, three colonies of the *P. pastoris* transformants were picked and cultured in 50 ml YPD medium at 30°C and 130 rpm on a shaker until reaching OD600 of around 10-14. The 50 ml cultures were then centrifuged first at 1000 xg for 5 min and then at 5000 xg for 8 min and the supernatant was 0.2 μ m sterile filtered.

2.6 Estimation of protein expression levels

Samples from all the culture filtrates were analysed by SDS-PAGE and the transformants with the strongest band of the desired size, or if the bands were similar the transformant with the most protein, were selected. From the selected culture 3 ml of culture filtrate were desalted on a Bio-Rad Econo-Pac® 10DG desalting column in accordance to the manual with buffer A (25 mM TRIS pH 7.5 and 200mM NaCl). An IMAC purification was performed to further purify the protein using a GE Healthcare HiTrap 1 ml Chelating HP charged with 0.1 M NiSO₄ and buffer A as running buffer. For the elution of the protein buffer B (25mM TRIS, 200mM

NaCl and 500mM imidazole) was used and the sample was eluted into 1 ml fractions.

The chromatography program differed somewhat for the different constructs. After loading the sample, the column was washed with 5 ml of buffer A. Elution was done step-wise with buffer B extracting the protein with 250 mM (Tr), 300 mM (α -factor) or 500 mM (Epx1, Kp+R and Kp+RPIKK) of imidazole. The flow-rate was 1 ml/min.

The Tr-sample fractions, which was the first sample to run through the IMAC, was visualized with an SDS-PAGE to see in which fraction the protein occurred. The fractions containing the protein was pooled together and an SDS-PAGE was performed.

3 Results

3.1 Design of gene constructs for expression

Since time was limited, the project had to be limited to five gene constructs where the native signal peptide (SP) of BaLOOS1 was exchanged with four different SPs. Their names and the origin of the SPs are listed in Table 1.

Table 1: Names of the signal peptide exchange gene constructs and origins of the signal peptides used.

Construct name	Original host	Original protein	Accession code
Tr	<i>T. reesei</i>	Loosenin-like protein (TrLOOL)	XP_006961781
Kp+R	<i>P. pastoris</i>	Loosenin-like protein (KpLOOL)	ANZ74303
Kp+RPIKK	<i>P. pastoris</i>	Loosenin-like protein (KpLOOL)	ANZ74303
Ex	<i>P. pastoris</i>	Extracellular protein X1 (Epx1)	F2QXH5
α -factor	<i>S. cerevisiae</i>	Mating factor	AJW01277

Two SPs were taken from the loosenin homolog from *Trichoderma reesei* (TrLOOL) and *P. pastoris* (KpLOOL), respectively, which were included in the previous pilot study. They were selected based on their seemingly high protein expression levels compared to BaLOOS1 with the native SP (Figure 1). The alpha-factor was chosen because it is the most common SP in commercially available plasmids for extracellular expression in *P. pastoris* and has successfully been used for expression of many different proteins. The alpha factor is 89 amino acids long, i.e. much longer than the other SPs and includes an additional enhancer element that is believed to promote the expression of the target protein, see Figure 3.

The forth SP is the Epx1-SP which comes from what is claimed to be the most abundant secretory protein of *P. pastoris* (Heiss, Puxbaum et al. 2015).

Prediction of SP cleavage sites using the Signal P 4.1 server revealed that the TrLOOL1 and BaLOOS1 proteins have similar cleavage sites. The SP ends with Ala followed by identical N-termini, Ala-Pro-Thr, indicating that cleavage at the same site is highly likely if their SPs are swapped. However, the corresponding region is different for KpLOOL, and the N-terminus of the mature protein may potentially affect cleavage recognition. Therefore, we set out to examine if it seems possible to replace any residues at the N-terminus of BaLOOS1 with corresponding residues from KpLOOL.

Four protein structures were retrieved from the PDB by BLAST search with the BaLOOS1 sequence as query; 3d30 is an expansin-like protein from *Bacillus subtilis*; 4jcw is the *Clavibacter michiganensis* expansin; 2hcz is a beta-expansin pollen allergen from maize; 5kjo is an endoglucanase from *Phanerochaete chrysosporium*. These structures were superposed in Pymol with a homology structure model of BaLOOS1, built using Phyre 2, to understand the structural importance of the N-terminus of BaLOOS1.

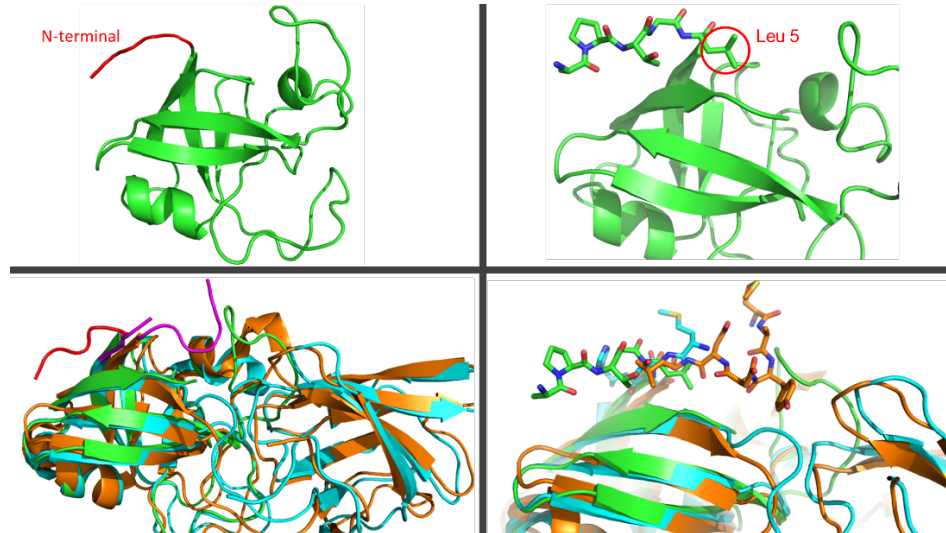


Figure 3: The top left panel shows the BaLOOS1 homology model in cartoon style with the N-terminus marked in red. The top right panel shows the N-terminus in sticks representation with the Leu 5 circled in red and the rest of the protein in cartoon. The left bottom panel shows the BaLOOS1 superposed with PDB: 3D30 in orange and 4JCW in teal while the BaLOOS1 N-terminus is red and the others magenta. In the bottom right panel, the same proteins are superposed, and the same colours are used but the N-terminals are shown in stick representation.

The comparison suggests that Leu 5 is the first amino acid at the N-terminus of BaLOOS1 with structural importance for the protein. The placement of the Leu 5 sidechain is suitable for hydrophobic interactions with nearby residues. Thus, it could be anchoring the N-terminus to the rest of the protein making it unwise to replace, see Figure 4. However, the preceding amino acids does not appear to have the same structural importance. Thus, making them ample candidates for a replacement corresponding to the N-terminal residues of the KpLOOL SP. This lead to two different constructs being designed with the KpLOOL SP. One only had the first N-terminal residue exchanged (Kp+R), while the other construct (Kp+RPIKK) had the first five residues of KpLOOL replace the first four residues of BaLOOS1 (Figure 4).

Alpha MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDAVLPPFSNSTN

BaLOOS1 MFFNAAIFVSVLAVVGFVIA APTGLQHGDGTFYATGLGACGKVNVDMDIA~
TrLOOL1 MVSFTLTALSAIAMALTTS APTGLQHGDGTFYATGLGACGKVNVDMDIA~
KpLOOL+R MQFLQTLTLAIISTFLSSAVG RPTGLQHGDGTFYATGLGACGKVNVDMDIA~
KpLOOL+RPIKK MQFLQTLTLAIISTFLSSAVG RPIKKLQHGDGTFYATGLGACGKVNVDMDIA~
EpxL-SP MKLSTNLILAIAAASAVVSA APTGLQHGDGTFYATGLGACGKVNVDMDIA~
Alpha NGLLFINTTIIASIAAKEEGVSLDKREAEA APTGLQHGDGTFYATGLGACGKVNVDMDIA~

Figure 4: Protein sequences of the signal peptides used in red, and the beginning of BaLOOS1 in black with the N-terminal changes in green. The alpha factor enhancer element is blue. The cleavage site between the signal peptides and the N-terminal is indicated by a gap in the sequence.

3.2 Gene assembly, DNA preparation and analysis

The plasmid from the previous pilot study was used as backbone since it already carried the GAP promotor and BaLOOS1 gene with the native SP with an appended C-terminal His₆-tag. Since the expression is under control of a constitutive GAP promotor the protein can be expressed with glucose as a carbon source. The plasmid also carries Ampicillin and Zeocin resistance markers for selection of positive transformants. However, there seemed to be something wrong with the backbone-PCR and several runs had to be conducted to attain the DNA amount needed for the transformation over to the *E. coli* Top10-strain.

The absorbance measurements of the *E. coli* plasmid preparations showed promising levels of plasmid DNA from both colonies, i.e. close to the desired 200 ng/μl, see Table 2. It is also seen in Figure 5 that the bands correlates reasonably with the size of the plasmid being (6740 bp), even though the DNA load was rather high.

Table 2: Spectrophotometric measurements of the two colonies picked from the *E. coli* cells containing the plasmid with the native BaLOOS1 DNA sequence.

Samples	Ng/μl	260/280	260/230
C1	230.9	1.90	2.35
C1	289.2	1.87	2.09
C2	274.5	1.89	2.19
C2	221.1	1.86	2.10

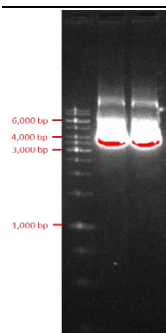


Figure 5: A cut out of the 1% agarose gel from the electrophoresis of the *E. coli* plasmid prep with the plasmid containing the BaLOOS1 native DNA sequence.

After the desired amount of plasmid DNA was reached the native BaLOOS1 SP was replaced by Gibson assembly, using synthesized DNA containing the respective SP sequences and flanking Gibson overlap regions, shown in Figure 2.

3.3 Transformation of *E. coli* and plasmid preparation

The Gibson assembly mixture was directly used for transformation of *E. coli* and subsequent Miniprep plasmid preparation, followed by PCR amplification of the region of interest and gel analysis of the amplified DNA for verification of the constructs. For all constructs, band sizes on the gel were in accordance with the expected (α -factor 928bp; Ex, 721bp; Kp+R, 724bp; Kp+RPIKK, 733bp; Tr, 721bp) as shown in Figure 6.

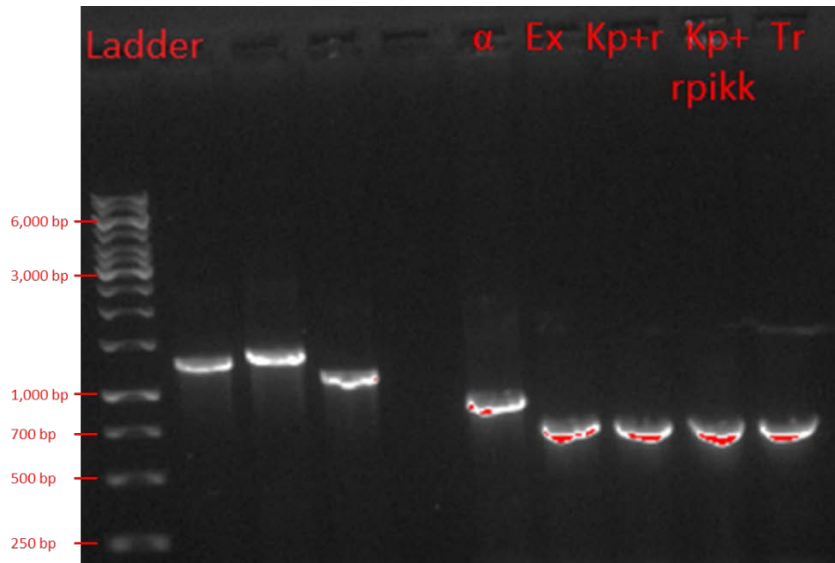


Figure 6: A 1% agarose gel of the DNA constructs with a ladder in lane 1, alpha-factor in lane 6, Epx1-SP in lane 7, Kp+R in lane 8, Kp+RPIKK in lane 9 and Tr in lane 10. The samples in the other lanes are not related to this project.

The Nanodrop absorbance measurements of the miniprep DNA from selected *E. coli* transformants showed very promising results. DNA concentrations were around 0.5-0.6 $\mu\text{g}/\mu\text{l}$, which is higher than the commonly desired 0.2 $\mu\text{g}/\mu\text{l}$ for transformation to *P. pastoris* (Table 3).

Table 3: Absorbance of the plasmid DNA from the *E. coli* transformants after DNA preparation.

Samples	ng/ μl	260/280	260/230
Tr	495.3	1.86	2.34
Kp+R	566.7	1.85	2.33
Kp+RPIKK	514.9	1.85	2.34
Ex	519.4	1.85	2.31
α -factor	566.0	1.85	2.30

After linearization of the plasmid and PCR clean up the Nanodrop measurements showed lower amounts of DNA in the samples but still judged to be sufficient for proceeding with *P. pastoris* transformation (Table 4).

Table 4: Absorbance of the plasmid DNA from *E. coli* after plasmid preparation, linearization and PCR clean-up, which shows some loss of DNA from the previous measurment.

Samples	ng/μl	260/280	260/230
Tr	323.1	1.84	2.23
Kp+R	304.1	1.83	1.96
Kp+RPIKK	328.6	1.82	1.60
Ex	284.8	1.83	2.01
α -factor	317.4	1.83	1.94

3.4 Transformation of *P. pastoris* and colony PCR analysis

After transforming the *P. pastoris* cells, a reasonable number of colonies appeared on the Zeocin plates for all constructs. Three colonies were picked for each of the constructs and the colony PCR, of which the gel can be seen in Figure 5. The gel shows the DNA of the BaLOOS1 constructs together with the upstream GAP promoter. The fragments have a size around 1000 kb with Tr being 1012bp; Ex being 1012bp; Kp+R being 1015bp; Kp+RPIKK being 1024bp; α -factor being 1219bp.

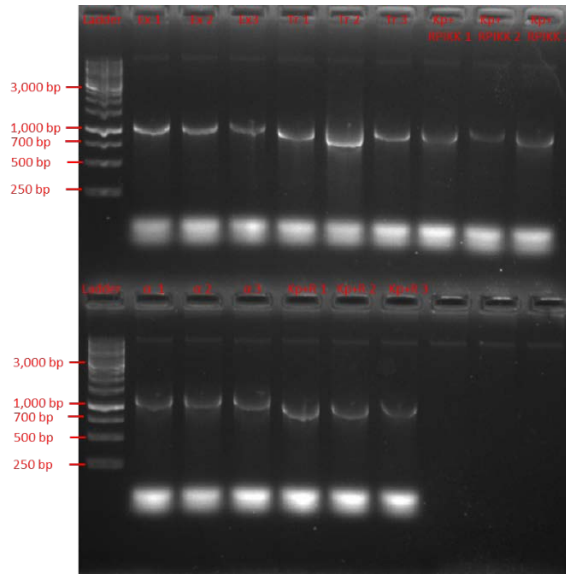


Figure 7: A picture of the 1% agarose gel with DNA from the cPCR of the *P. pastoris* transformants with three different transformants of each construct. With the loading in the top lanes as follows; ladder in lane 1, Ex in lanes 2-4, Tr in lanes 5-7, Kp+RPIKK in lanes 8-10. Loading in the bottom lanes as follows: Ladder in lane 1, alpha-factor in lanes 2-4, Kp+R in lanes 5-7. All of the transformants have bands of the desired size.

The difference seen between colonies of the same construct is most likely due to a difference in the DNA loading amount, which seems reasonable when comparing band intensity and location on the gel.

3.5 Protein expression

After the 3-day growth, OD600 of the culture was measured for one of the colonies from each construct. The results are shown in Table 5.

Table 5: Absorbance of the 50 ml *P. pastoris* cultures that were grown to express the BaLOOS1 protein. Only one transformant from each construct was measured since they seemed to be close in OD.

Samples	OD600
Tr	14.3
Kp+R	9.8
Kp+RPIKK	10.0
Ex	10.0
α -factor	11.0

After filtration through a 0.2 μ m filter, the supernatant was analysed with SDS-PAGE to decide which of the transformants from each construct had the highest protein expression, see Figure 8.

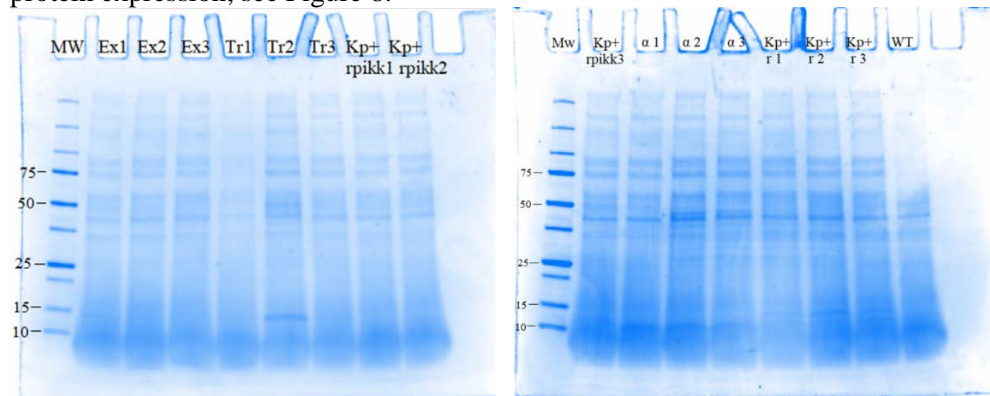


Figure 8: The figure shows the two SDS-PAGE gels from the run of the 0.2 μ m filtered culture filtrates. The left gel is loaded with the molecular weight marker (MW) in lane 1, Ex in lanes 2-4, Tr in lanes 5-7, Kp+RPIKK in lanes 8-9. The right gel has the MW in lane 1, Kp+RPIKK in lane 2, alpha-factor in lanes 3-5, Kp+R in lanes 6-8, Wild type (WT) in lane 9. The BaLOOS1 can be clearly or slightly seen in all transformants at 11.4 kDa with the Tr2 having the strongest band.

The BaLOOS1 is calculated to be 11.4 kDa with the help of ProtParam from ExPASy (<https://web.expasy.org>). Within the expected size range, it is seen that the Tr2 transformant has the strongest band on the gel. The correct band size can also be seen in the alpha-factor and Kp+R, although only slightly. For Kp+R, transformant 2 also seems to have the strongest band. For the alpha factor transformant 2 seemed to have the highest overall protein expression. For the Ex and Kp+RPIKK constructs the transformants had very similar protein expression levels, therefore the transformant 2 was chosen for simplicity.

Purification of the BaLOOS1 was done by trapping the protein on an IMAC column and eluted it with imidazole into several fractions. For the Tr construct, individual fractions from the IMAC purification were analysed by SDS-PAGE. Fractions 2 and 3 gave strong bands in the area of the expected protein while faint bands could be seen in the adjacent fractions. It is also noticed that the Tr flow-through does not seem to contain any protein of the desired size (Figure 9).

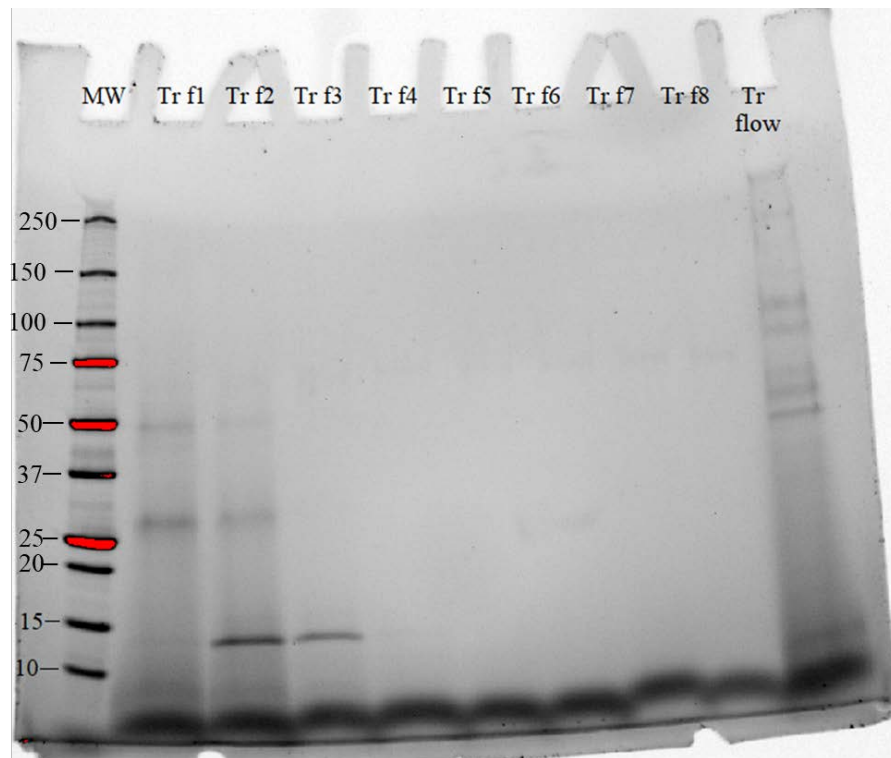


Figure 9: A depiction of the SDS-PAGE gel with the fractions of the Tr construct after the IMAC purification to see in which fractions the protein resides. Fractions 4-11 and the flow-through were analysed. The loading was with the molecular weight marker (MW) in lane 1, the Tr fractions 4-11 in lanes 2-9 followed by the flow-through in lane 10. The band for the BaLOOS1 at 11.4 kDa can clearly be seen in lanes 3-4 with slight bands in the adjacent lanes 2 and 5. No band for BaLOOS1 can be seen in the flow-through.

For all constructs, including Tr, all the fractions that were expected to contain BaLOOS1 were pooled together for each sample to quantitatively recover as much as possible of the BaLOOS1 that was originally present in the culture filtrate. The pooled protein was then analysed by SDS-PAGE to compare the difference in BaLOOS1 content, see Figure 10.

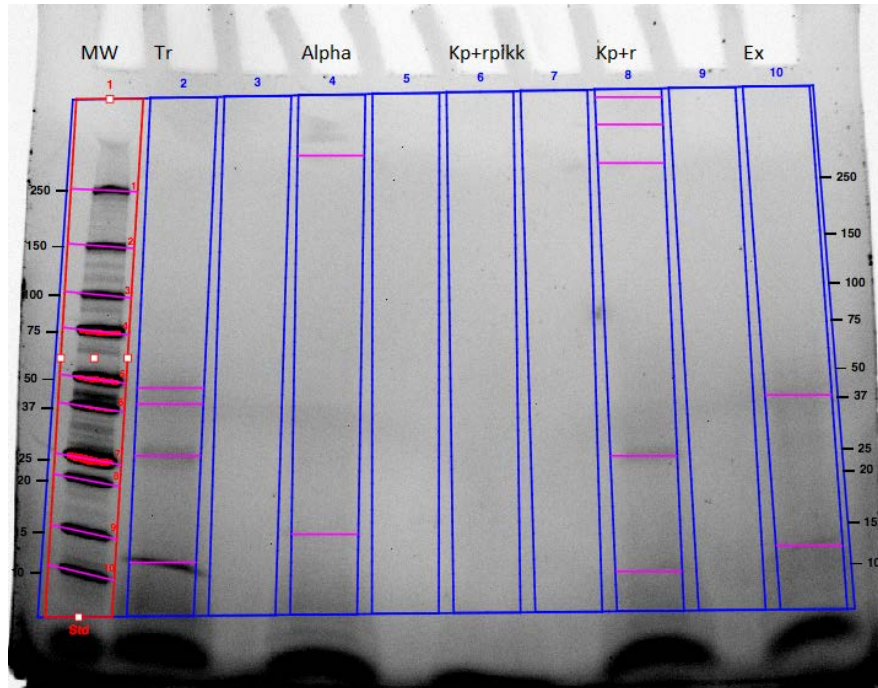


Figure 10: The figure depicts the SDS-PAGE gel after being processed by the Bio-Rad gel imaging software. The loading was with the molecular marker (MW) in lane 1, Tr in lane 2, alpha-factor in lane 4, Kp+RPIKK in lane 6, Kp+R in lane 8, Ex in lane 10. The purple bands are indicating where in the lanes there is protein. The transformants with a band in the size of BaLOOS1 11.4 kDa were the Tr, Kp+R and the Ex constructs.

The Tr construct shows the strongest band, by far, of the expected size, compared to the others. This was also confirmed by using the Bio Rad Image Lab 6.0 which calculated the relative amounts of protein along the lanes. For Tr, the proportion of BaLOOS1 was estimated at ~80% of the pooled protein from the IMAC column.

A rough estimation of the protein content for each of the pooled BaLOOS1 constructs was done by observation of the gel in Figure 10 and concluded that the Tr construct had the highest protein expression of BaLOOS1 (Table 6).

Table 6: The table compares the expression of the BaLOOS1 between the different constructs.

Samples	Estimated expression level
Tr	+++++
Kp+R	+
Kp+RPIKK	-
Ex	+++
α -factor	++

4 Discussion

Since the aim of the study was to exchange the native SP for BaLOOS1 in an attempt to increase the expression level, it is deemed a success. All the different constructs were successfully inserted, and the *Trichoderma* SP seemed to give the highest expression. The Tr-SP expression might be enough for continued production and for studies of structure and function.

The reason that the Tr-SP gave the highest expression level could be due to the similarity to the BaLOOS1 SPs sequence, cleavage site and the N-terminus. The similar overall protein sequences could indicate a shared ancestor and would be interesting for further studies.

However, the higher expression levels could as well be due to a more favourable placement within the genome of the transformants. It would also be of interest to make a comparison of expression levels with the native BaLOOS1 SP to see if the expression level of the Tr-construct is higher.

The α -factor signal peptide is noticeably longer than all the other signal peptides used in this project. The reason for this is that it is containing an enhancer element which has shown to give a high protein expression in *P. pastoris*. This has also lead to the α -factor becoming the most common signal peptide sold for commercial use. However, since the original protein sequence of the N-terminus associated with the α -factor was not considered in this project it might not be optimal for the BaLOOS1 protein.

The Epx1-SP was constructed based on the Epx1 secretory leader which is the most abundant secretory protein of *P. pastoris* (Heiss, Puxbaum et al. 2015). This is not a commonly used peptide but since the cleavage site was similar to the BaLOOS1 it was still an interesting peptide. The reason for the poor protein expression for this protein could be due to the transformants. However, it is also possible that since the original protein has a pro-sequence the lack of that pro-sequence in the used construct may have affected the expression (Heiss, Puxbaum et al. 2015). This could be tested in coming studies.

The Kp constructs are based on the *P. pastoris* loosenin-like protein and is a protein which have had a good expression in the pilot study (Figure 1). However, for the two constructs created with this SP the N-terminus of BaLOOS1 was altered since the BaLOOS1 N-terminal sequence and the KpLOOL did not match. The SP and the N-terminus of the constructs Kp+R and Kp+RPIKK were created to investigate the effect of amino acid change in the N-terminal region. However, in hindsight the experiment could have been more informative if a Kp-SP construct without N-terminal modification had been included in the study, and perhaps the Kp+RPIKK construct would have been more effective with only one Lys added instead of two since that would probably not affect the cleavage site.

The transformation of the *E. coli* Top10 competent cells were performed several times because of difficulties to get enough plasmid DNA from the backbone PCR. After trying many ways to alter the PCR program, changing machine and changing reagent ratios it was suspected to be a defective batch of dNTPs. Fortunately, after several attempts the samples acquired could be pooled together to get a high enough DNA content to transform the cells successfully. Later, the selection of transformants failed, this was due to an old batch of ampicillin and the problem was remedied after using a new batch of ampicillin. These problems demonstrate the importance of high quality products and to work with freshly synthesized stock solutions.

It is also worth mentioning that the Phyre 2 model used for the Pymol super positioning and homology study is only a prediction which has limitations. This further incentivise to structurally determine the protein so that accuracy for coming experiments can be heightened.

It is desirable to estimate and compare the expression levels for all constructs, preferably also with the native SP. However, the data needed to calculate the expression was more difficult than expected to obtain. This could be due to many reasons with the most likely being the presence of imidazole interfering with the measurement of absorbance, the BaLOOS1 is not the only protein in the solution and the calculated extinction coefficient for BaLOOS1 is low. Since the project had limited time, these issues could not be resolved. Although, additional attempts to measure the absorbance of purified BaLOOS1 has been done after the project ended. These attempts have also failed to get any readings at 280 nm, even though a clear band with the correct size have been seen on an SDS-PAGE gel.

Before further studies of BaLOOS1 it would be wise to confirm that the protein band seen on gel is actually BaLOOS1, e.g. by peptide mapping. For the coming upscaling it would be a good idea to also grow the cells with the BaLOOS1 native SP to be able to do a side-by-side comparison of expression levels with the TrLOOL-SP variant. In principle it should be possible to pool together the protein from the two different batches for a higher protein content since they should be identical after the signal peptide is cleaved off. Also, glycosylation is not expected to vary between batches, since there are no N-glycosylation sites in the BaLOOS1 sequence. To make sure the protein is correctly folded and has maintained its activity, it would also be wise to perform an activity study. A suggestion for that could be to treat cotton fibres or filter paper with a solution containing BaLOOS1 and observe it in a microscope to see if the cellulose microfibrils are disrupted or appear swollen. Another interesting test would be to study the viscosity of a BaLOOS1 treated cellulose solution to see if the cellulose is separated enough to get thinner fibres and lower viscosity, or if swelling of the cellulose rather increases the viscosity.

References

- Berman, H. M., et al. (2000). "The Protein Data Bank." Nucleic Acids Research **28**(1): 235-242.
- Castillo, R. M., et al. (1999). "A six-stranded double-psi β barrel is shared by several protein superfamilies." Structure **7**(2): 227-236.
- Dal Piccoli, T., et al. (2018). "High-performance of *Agaricus blazei* fungus for the biological pretreatment of elephant grass." Biotechnology Progress **34**(1): 42-50.
- Edgar, R. C. (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." Bmc Bioinformatics **5**: 1-19.
- Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.
- Galicía-Medina, C. M., et al. (2018). "Current state of bioethanol fuel blends in Mexico." Biofuels Bioproducts & Biorefining-Biofpr **12**(3): 338-347.
- Heiss, S., et al. (2015). "Multistep processing of the secretion leader of the extracellular protein Epx1 in *Pichia pastoris* and implications for protein localization." Microbiology-Sgm **161**: 1356-1368.
- Kapp, K., et al. (2013). "Post-targeting functions of signal peptides."
- Kelley, L. A., et al. (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nature Protocols **10**(6): 845-858.
- Kumar, S., et al. (2016). "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." Molecular Biology and Evolution **33**(7): 1870-1874.
- Lopez, S. C., et al. (2018). "Induction of Genes Encoding Plant Cell Wall-Degrading Carbohydrate-Active Enzymes by Lignocellulose-Derived Monosaccharides and Cellobiose in the White-Rot Fungus *Dichomitus squalens*." Applied and Environmental Microbiology **84**(11): 11.
- Loyd, A. L., et al. (2018). "Elucidating wood decomposition by four species of *Ganoderma* from the United States." Fungal Biology **122**(4): 254-263.
- Martínez-Gutiérrez, E. (2018). "Biogas production from different lignocellulosic biomass sources: advances and perspectives." 3 Biotech **8**(5): 18.
- McQueenmason, S. and D. J. Cosgrove (1994). "DISRUPTION OF HYDROGEN-BONDING BETWEEN PLANT-CELL WALL POLYMERS BY PROTEINS THAT INDUCE WALL EXTENSION." Proceedings of the National Academy of Sciences of the United States of America **91**(14): 6574-6578.

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. Protein Function Prediction: Methods and Protocols. D. Kihara. Totowa, Humana Press Inc. **1611**: 59-73.

Payne, C. M., et al. (2015). "Fungal Cellulases." Chemical Reviews **115**(3): 1308-1448.

Quiroz-Castaneda, R. E., et al. (2011). "Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*." Microbial Cell Factories **10**: 9.

Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Suzuki, H., et al. (2014). "Sequence diversity and gene expression analyses of expansin-related proteins in the white-rot basidiomycete, *Phanerochaete carnosae*." Fungal Genetics and Biology **72**: 115-123.

Wang, Y. X., et al. (2003). "Manganese-lignin peroxidase hybrid from *Bjerkandera adusta* oxidizes polycyclic aromatic hydrocarbons more actively in the absence of manganese." Canadian Journal of Microbiology **49**(11): 675-682.

Acknowledgements

I want to thank my supervisor Jerry Ståhlberg for always, happily, sharing his knowledge with me and putting up with the endless stream of questions that I have bombarded him with.

I also want to thank Topi Haataja for his patience during the late hours in the lab and all my untimely interruptions of his work.